图书情报工作
第66卷 第7期 2022年4月

景点文化资源标签自动生成与应用研究*

■ 郑淞尹 谈国新
华中师范大学国家文化产业研究中心 武汉 430079

摘 要:[目的/意义]为旅游景点生成高质量的文化资源标签,解决文化旅游服务中信息检索困难、推荐形式单一的问题。[方法/过程]首先,设计包含显式和隐式两种标签类型的文化资源标签体系;然后,提出基于特征词筛选和噪声词过滤的显式标签生成方法,以及设计隐式标签中文化感知强度和文化感知相似度的计算方法,并基于以上方法生成景点文化资源标签;最后,针对旅游信息服务中的不同场景,提出基于文化资源标签的检索和推荐方法。[结果/结论]以武汉市的真实旅游数据为例进行实证研究,结果表明,基于本文方法生成的标签能够准确刻画景点的文化资源特征,基于标签的检索和推荐方法均具备较强的可解释性,可有效提升信息服务的透明度和用户对结果的信任度,对其他领域的推荐解释性研究具有参考价值。

关键词:旅游信息服务 标签生成 资源检索 旅游推荐
分类号:G254
DOI:10.13266/j.issn.0252-3116.2022.07.013

1 引言

文化旅游是以文化资源为主要依托的旅游活动,因其能更好地满足旅游者日益增长的精神文化需求,正受到越来越多旅游者的喜爱[1]。景点是文化的载体,蕴含着丰富的文化资源,但景点文化资源的多样性和用户文化偏好的差异性使得用户在出行前往往需要查询大量信息,才能找到感兴趣的景点。调查发现,知名度和用户数都排名前列的在线旅游服务平台如携程、马蜂窝、猫途鹰等,只包含有少量大众化的文化旅游相关信息,如携程网提供了以“人文旅游”为主题的旅游产品,猫途鹰对“历史遗迹”“圣地与宗教”等类型的景点进行了聚类,但都没有提供专门针对文化旅游信息的检索和推荐服务,用户正面临着信息检索困难,决策效率低下的问题。因此,如何充分挖掘和准确描述景点的文化资源,为具有不同文化偏好的用户提供精准化的检索和推荐服务,是当前亟需解决的问题。

标签是指能够反映资源关键特征的词或短语,因其能够有效提升资源组织管理的效率,受到了学界和业界的共同关注[2-3]。国内具有代表性的社会化网站如豆瓣、知乎、微博等,标签在其资源分类、信息检索和个性化推荐中发挥着十分重要的作用[4]。标签易于理解和便于处理的特性使其同样适用于对旅游资源的组织管理,如果标签能够全面、准确地描述景点的文化资源,那么用户就可以根据标签内容快速获知景点的文化特征,在线旅游平台也可以基于标签提供更好的检索和推荐服务。移动互联网和在线旅游的快速发展使得网络上产生了大量在线旅游信息,如旅行游记、用户游览历史、用户评论等,这些信息不仅真实反映了用户的旅游体验,而且从不同角度对景点的资源和服务进行了描述和评价。因此,笔者将通过挖掘在线旅游信息,生成质量较高的景点文化资源标签,为文化旅游信息的检索和推荐服务奠定基础。

2 相关研究

本文旨在通过挖掘在线旅游信息,自动生成可应用于文化旅游信息服务的景点标签。因此,本文的研究内容主要涉及标签自动生成与景点标签两个方面,本节将从这两个方面对相关研究进行梳理和总结。

2.1 标签自动生成研究

标签生成是标签应用的前提,目前标签生成主要有3种方式:①用户标注;②专家或管理者标注;③算

* 本文系文化和旅游行业标准化研究项目“乡村文化资源分类体系研究”(项目编号:WH/Y 09-2019)和国家文化和旅游科技创新工程项目“湖北省非物质文化遗产数字化传播创新平台研发”(项目编号:2019-008)研究成果之一。
作者简介:郑淞尹,博士研究生;谈国新,教授,博士生导师,通信作者,E-mail:gxtan@mail.ccnu.edu.cn。
收稿日期:2021-10-29 修回日期:2022-01-16 本文起止页码:132-143 本文责任编辑:杜杏叶

法自动生成[5]。社会化网站能够拥有大量标签得益于用户群体的自发标注，然而部分在线平台虽然也拥有一定规模的用户，但由于没有开放用户标注功能或开放时间较晚，导致没有标签或标签数量不足。此外，随着大数据时代的到来，很多行业的数据量都急剧增长，数据更新速度越来越快，这使得由专家或管理者人工添加标签的方式存在一定的操作性和时效性问题。因此，如何在尽可能减少人工参与的情况下，通过算法自动生成高质量标签，成为了研究的重点。

目前已有学者在视频[6]、医疗[5,7]、社交[8-10]、知识服务[11-12]等多个领域开展了相关研究，并提出了适用于领域数据的标签自动生成方法。Z. Shen 等[6]提出了基于传感器元数据的户外视频自动标注方法，该方法首先将视频中的场景建模为几何形状，然后通过地理信息数据库查询几何形状对应的地理对象，并提取其文本信息作为视频标签；孟晴秋等[5]基于医生在线问诊文本信息，提出了结合时间周期特征与文本主题特征的医生标签自动生成算法；吴小兰等[8]提出了结合用户关系网和标签共现网的微博用户标签预测方法，该方法首先使用带重启的随机游走模型生成候选标签，然后基于标签链抽取候选标签推荐给目标用户；熊回香等[9]提出了基于 LDA 主题模型的微博标签生成方法，该方法首先根据主题模型和用户微博生成预选标签，然后通过分析用户关注的人及用户粉丝的微博对预选标签进行控制，生成最终标签；L. Zeng 等[11]结合软件工程经验知识和深度学习算法，提出了一种代码标签生成方法；赵辉等[12]以科技管理部门的情报需求为研究对象，利用主题词抽取、TF-IDF 等自然语言处理算法生成特征标签，并利用协同过滤、标签关联等推荐算法为具有相似特征的科技管理组织提供内容推荐。此外，对标签生成质量的评估也是研究的重点，李蕾等[13]比较分析了多种标签质量评估方法，并将已有方法划分为人工评估、基于标签自身统计属性、依据规范词语等 8 种不同类型。章成志等以图片标签[14]、博文标签[15-16]为对象开展了系列研究，提出标签的社会化属性可以作为区分标签质量的重要特征，并通过融合标签内容属性和社会化属性训练得到了具有更高性能的标签质量自动评估模型。

2.2 景点标签研究

虽然基于信息挖掘的标签自动生成在以上领域已取得了一定成果，但旅游领域中相关研究却很少，且研究重心均是标签应用而非标签生成。例如，李雅美等[17]提出了基于地域、时间、主题、类型等特征标签的景点推荐方法，但使用的标签均由人工生成；史一帆等[18]通过引入景点类型标签信息，提出了基于用户社会关系和景点标签的协同过滤推荐算法，实验表明该方法具有更高的推荐准确率。但文中使用的景点类型标签同样是由人工生成，且标签内容仅包含了"湖泊、河流""山岳、山岭"等对景点的高度概述，无法通过该标签获知更多与景点相关的信息。此外，还发现了少量与游客画像、景点实体识别相关，同时也涉及标签生成的研究[19-23]。例如，单晓红等[19]以携程网的酒店评论为基础，通过抽取用户信息属性、酒店信息属性和评价信息属性，构建了用户画像和酒店特征画像；刘海鸥等[20]基于游客基本信息标签、行为信息标签与情境信息标签，结合本体方法构建了游客用户画像，并提出了基于用户画像的旅游情境化推荐模型。

梳理文献发现，已有与景点标签相关的研究在标签生成和标签应用上都存在一定不足。首先，在景点标签生成方法上，已有研究大多局限于使用人工的方式获取所需的景点标签，这不仅需要耗费大量人力，还可能因为人力不足或时间不够，导致生成的标签存在数量较少、粒度较粗、内容较单一等问题，且标签更新成本很高。其次，在景点标签生成视角上，已有研究虽然从不同维度生成了景点标签，但对景点文化资源信息进行描述的标签很少，且目前尚未发现较为完善的景点文化资源标签体系。再次，在景点标签应用上，已有研究主要关注于如何基于标签进行景点推荐，而对景点检索方法的研究较少。针对上述问题，笔者从文化旅游视角出发，设计景点文化资源标签体系，提出景点文化资源标签自动生成方法，通过对在线旅游信息的挖掘，实现在只有少量人工参与的情况下，自动生成可应用于文化旅游信息检索和推荐的高质量标签。

3 研究框架与关键步骤

3.1 研究框架

本研究包括 3 个子任务：①设计景点文化资源标签体系；②设计标签生成算法并生成各维度标签；③分析标签应用场景，展示具体应用实例。针对以上任务，构建了图 1 所示的研究框架，包括标签体系设计、数据采集与预处理、标签生成及可视化、标签应用 4 个步骤。首先，以文化层次理论为基础，结合旅游信息服务中用户的实际需求，设计包含显式和隐式两种类型的景点文化资源标签体系；其次，从在线旅游服务平台采集景点文化主题数据和景点评论语料，并对数据进行清洗、转换、分词、去停用词等处理；然后，设计标签生

成算法，生成各景点的文化资源标签，并通过词云图和数据表对标签进行可视化展示；最后，以景点的检索和推荐为例，分析文化资源标签在旅游服务场景中的具体应用。

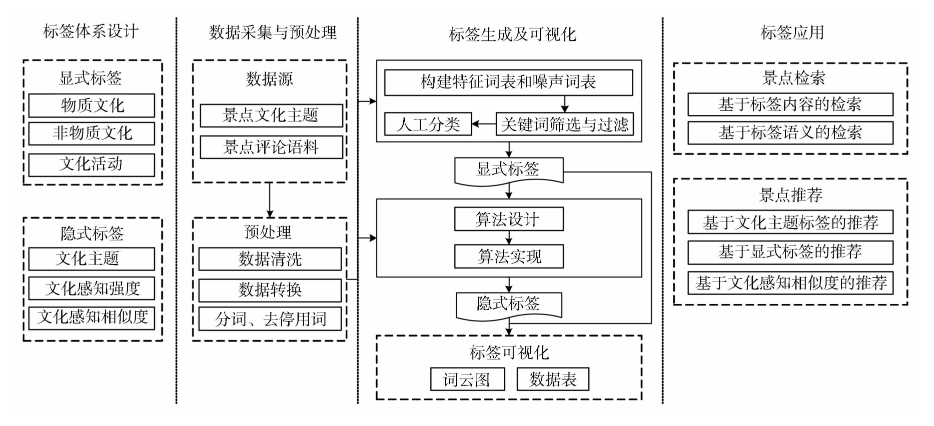


图 1　研究框架

3.2　关键步骤

3.2.1　标签体系设计

标签体系的设计综合考虑了 4 个方面因素：①文化的构成维度；②已有研究对文化感知维度的划分[24-25]；③景点文化资源自身特点；④用户对文化旅游信息服务的实际需求。标签体系设计的具体过程和依据如下：首先，文化层次理论认为文化由 3 个维度构成，分别是物质文化、制度文化和心理文化[26]。在已有文化感知的相关研究中，尹小娜等[24]将游客文化感知分为物质文化感知和非物质文化感知；李东晔等[25]以文化层次理论为基础，将游客文化感知分为文化氛围感知和文化产品感知，其中文化产品感知的主要内容之一是对文化活动的感知。因此基于已有研究成果，设计了物质文化标签、非物质文化标签和文化活动标签。其次，文化主题能够有效反映景点文化资源的类型特征，一方面旅游门户网站可以对具有相同文化主题的景点进行聚类和展示，另一方面用户可以根据自己的文化偏好快速查找相应主题的景点，因此在标签体系中加入了文化主题标签。再次，用户在进行旅游决策时，不仅需要知道景点是否包含了自己感兴趣的文化资源，同时也会关注其他用户对该文化资源进行感知体验的实际情况，因此设计了文化感知强度标签，将用户对文化资源的感知程度进行量化表示。最后，为用户推荐相似景点的前提是获知各景点在文化特征上的相似度，因此设计了文化感知相似度标签，当已知用户对某个景点感兴趣时，可以基于景点间的相似度为其推荐相似景点。

笔者按生成方式的不同将上述标签分为显式和隐式两类。其中，显式标签指可利用 TF-IDF[27]、TextRank[28] 等算法，直接从文本中抽取的关键词或短语，如“青花瓷、越王勾践剑”。隐式标签指需要对显式数据进行统计分析后才能得到的信息，如文化感知强度、文化感知相似度。此外，由于文化主题标签同样无法直接从文本中抽取得到，因此将文化主题标签也划分为隐式标签，标签体系具体内容如表 1 所示：

表 1　景点文化资源标签体系

标签类型	标签内容	标签概念及生成方法
显式标签	物质文化标签	用于描述景点物质文化资源的文本型标签，利用算法从评论数据中抽取
	非物质文化标签	用于描述景点非物质文化资源的文本型标签，利用算法从评论数据中抽取
	文化活动标签	用于描述景点文化活动资源的文本型标签，利用算法从评论数据中抽取
隐式标签	文化主题标签	用于描述景点文化资源类型特征的文本型标签，从在线旅游服务平台“望路者旅游网”获取
	文化感知强度标签	用于描述用户对景点文化资源认知和体验程度的数值型标签，通过挖掘显式标签数据获取。用 1－10 表示，数值越大，强度越强。文化感知强度包括 3 个指标：①整体文化感知强度；②基于文化主题标签的文化感知强度；③基于显式标签的文化感知强度
	文化感知相似度标签	用于描述用户对不同景点的文化资源产生相似感知的程度，通过挖掘显式标签数据获取。用归一化后的 0－1 表示，数值越大，相似度越高

3.2.2　标签生成方法

（1）显式标签生成。景点的文化资源可分为物质文化、非物质文化和文化活动 3 类，景点文化资源既是用户进行旅游决策的重要依据，也是用户在旅游过程

中感知体验的主要对象。用户评论可以真实反映用户对景点中各类文化资源的感知情况，评论中的高频要素反映了用户共同关注的文化吸引物。因此，笔者通过提取评论中不同词性的高频词和高频短语生成显式标签，具体流程如图 2 所示：

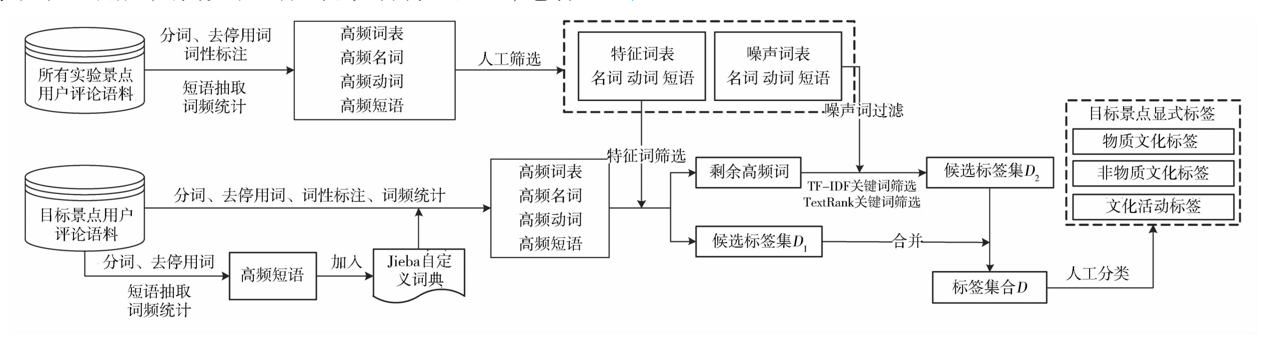


图 2　显式标签生成流程

显式标签生成包括 3 个主要步骤，具体如下：①构建特征词表和噪声词表。将所有实验景点的评论语料合并为一个文档，对其进行分词、去停用词、词性标注和短语抽取等处理，其中分词和词性标注使用 Jieba 工具[29]，去停用词使用哈工大停用词表，短语抽取指合并相邻词汇获取 2-gram 短语。由于物质文化和非物质文化标签多为名词，文化活动标签多为动词，因此提取词频最高的前 N_n 个名词、前 N_v 个动词和前 N_p 个短语构成高频词表。然后从高频词表中人工选取符合要求的词汇和短语作为特征词，存入特征词表，剩下的所有词作为噪声词，存入噪声词表。②提取目标景点标签。首先，对目标景点的评论语料进行预处理，将提取的高频短语加入 Jieba 自定义词典，其目的在于利用 TF-IDF、TextRank 提取关键词时，短语能够作为一个整体被提取出来。然后重新对评论语料进行预处理，提取词频最高的前 T_n 个名词、前 T_v 个动词和前 T_p 个短语构成目标景点的高频词表。其次，使用特征词表对高频词进行筛选，得到目标景点的候选标签集 D_1 和剩余高频词。然后，分别使用算法从评论语料中提取得分最高的前 N 个关键词，得到关键词表 K_1、K_2，依次使用噪声词表和关键词表 K_1、K_2 对剩余高频词进行过滤和筛选，得到候选标签集 D_2。最后，合并 D_1、D_2，得到目标景点的显式标签集合 D。③标签分类。对集合 D 进行人工分类，得到物质文化标签、非物质文化标签和文化活动标签。

（2）文化主题标签生成。望路者旅游网（whlyw.net）是一个旅游信息服务平台，平台通过人工的方式对景点的文化主题进行了标注，并以市级行政区为单位，按文化主题的不同对景点进行聚类。以武汉市为例，平台搜集了武汉市 69 个文化旅游景点的相关信息，并将其聚类到以生态文化、建筑文化、名人文化为主的 22 个不同文化主题下，同一个景点根据文化资源的特征，既可以只归属于一个主题，也可以同时归属于多个主题。笔者从望路者旅游网采集研究所需的文化主题信息，并将采集到的“文化主题—景点”矩阵转换为“景点—文化主题”矩阵，得到每一个景点对应的文化主题标签。

（3）文化感知强度标签生成（cultural perception intensity，CPI）。文化感知强度指用户在实际旅游中对景点文化资源的认知和体验程度，认知越深、体验越丰富，获得的文化感知越强。笔者对文化感知强度进行量化，设计了整体文化感知强度、基于文化主题标签的文化感知强度和基于显式标签的文化感知强度 3 个指标，计算方法如下：

- 整体文化感知强度。以景点为计算对象，每一个景点对应一个整体文化感知强度。用户的感知强度会受到物质文化、非物质文化和文化活动的共同影响，用户在评论中提及的文化资源对象越多，文化属性词汇出现的频率越高，说明其获得的感知越强。因此，笔者将所有显式标签词频数的加权平均定义为整体文化感知强度指标，如公式（1）所示：

$$CPI(S_k) = \frac{\sum_{i=1}^{3}\sum_{j=1}^{m_i} w_i * n_{ij}}{N} \qquad \text{公式(1)}$$

其中，$CPI(S_k)$ 表示景点 S_k 的整体文化感知强度，数值 3 表示 3 种不同类别的显式标签，m_i 表示第 i 类标签包含的标签个数，w_i 表示第 i 类标签的权重，n_{ij} 表示第 i 类标签中第 j 个标签的词频数，N 表示目标景点的评论总数。

- 基于文化主题标签的文化感知强度。以文化主题标签中的不同主题为计算对象，每一个主题对应一个文化感知强度。计算该指标有两个目的：一是比

较用户对同一景点的不同文化主题的感知差异,例如景点“东湖”包含“生态文化、楚文化”两种文化主题,但用户对“生态文化”的感知强度明显高于“楚文化”;二是比较用户对不同景点的同一文化主题的感知差异,例如“黄鹤楼”和“毛泽东同志旧居”都包含“建筑文化、名人文化”两种文化主题,但用户对黄鹤楼“建筑文化”的感知强度明显高于毛泽东同志旧居,而对毛泽东同志旧居“名人文化”的感知强度明显高于黄鹤楼。由于以上差异的存在,笔者对不同主题的文化感知强度进行量化表示,考虑到主题所包含的标签的数量和词频与感知强度正相关,因此设计了以下计算方法,如公式(2)所示:

$$CPI(T_i) = \frac{\sum_{j=1}^{m} n_j}{N} \qquad \text{公式 (2)}$$

其中,$CPI(T_i)$表示主题 T_i 的文化感知强度,m 表示主题 T_i 包含的标签个数,n_j 表示第 j 个标签的词频数,N 表示目标景点的评论总数。

- 基于显式标签的文化感知强度。以不同类别的显式标签为计算对象,每一类显式标签对应一个文化感知强度。计算该指标的目的在于比较用户对不同景点的同一类显式标签的感知差异,例如“恩施土家女儿城”有摆手舞、女儿会、哭嫁等丰富的民俗表演,用户对其“文化活动”的感知强度会明显高于“黄鹤楼”“东湖”等景点。因此,在计算得到每一个景点包含的各类显式标签的文化感知强度后,便可以根据用户偏好提供推荐。例如用户的需求是寻找“文化活动”丰富的景点,就可以根据“文化活动”标签感知强度的高低对景点进行排序,得到推荐列表。计算方法如公式(3)所示:

$$CPI(E_i) = \frac{\sum_{j=1}^{m} n_j}{N} \qquad \text{公式(3)}$$

其中,$CPI(E_i)$表示 E_i 类标签的文化感知强度,m 表示 E_i 类标签包含的标签个数,n_j 表示第 j 个标签的词频数,N 表示目标景点的评论总数。

(4)文化感知相似度标签生成(cultural perception similarity, CPS)。文化感知相似度指用户在旅游时对不同景点的文化资源产生相似感知的程度。由于显式标签是基于用户感知结果生成的对景点文化资源的描述,所以不同景点的显式标签在语义上越相近,用户在景点产生的文化感知就越相似。因此,笔者设计了一种基于标签语义相关性的文化感知相似度计算方法,具体包括以下 3 个步骤:

其一,获取标签词向量。基于 Word2Vec[30-31] 训练生成的词向量能够有效表示词语之间的语义关系,并可以通过计算词向量的余弦距离度量词与词的相似度。笔者以所有景点的评论数据为语料库,使用 Python 的 Gensim 工具包进行 Word2Vec 词向量训练,训练算法选择 Skip_gram,词向量维度设置为 128,训练得到每个标签的向量表示。

其二,计算文化感知特征向量。显式标签反映了用户文化感知的具体内容,文化感知特征向量是对文化感知内容的向量化表示,笔者借鉴文本向量化的方法,对所有显示标签的词向量进行加权平均,其中标签的权值等于经过标准化处理后的 TF 值。特征向量的计算公式如下:

$$Vec(S_k) = \frac{\sum_{i=1}^{n} tf_i * vec_i}{n} \qquad \text{公式(4)}$$

其中,$Vec(S_k)$表示景点 S_k 的文化感知特征向量,tf_i 表示第 i 个标签的 TF 值,vec_i 表示第 i 个标签的词向量,n 表示标签总数。

其三,计算景点文化感知相似度。余弦相似度常被用于评估向量之间的相似性,因此笔者通过计算特征向量的余弦值来评估景点之间的文化感知相似度,计算公式如下:

$$CPS(S_i, S_j) = \frac{\sum_{k=1}^{n} Vec(S_i)_k \times Vec(S_j)_k}{\sqrt{\sum_{k=1}^{n} (Vec(S_i)_k)^2 \times \sum_{k=1}^{n} (Vec(S_j)_k)^2}} \qquad \text{公式(5)}$$

首先根据上述方法构建景点文化感知相似度矩阵,然后通过查找矩阵,就可以获取任意景点之间的文化感知相似度。

4 实验与结果分析

本节将使用真实的旅游数据对所提方法进行验证,具体包括显式标签生成及评估、隐式标签生成及评估、标签展示、基于标签的景点检索以及基于标签的景点推荐等内容。

4.1 数据采集与预处理

以武汉市的文化旅游景点为研究对象,使用 Python 语言编写网络爬虫程序,从携程网(ctrip. com)采集了评论数较多的 57 个景点下 34 785 位用户的 57 324条评论数据,每条数据都包含景点名称、用户 ID、评论内容、评分和评论时间等信息,采集时间截至 2021 年 3 月。对数据进行以下预处理:①剔除内容为空和非汉字的评论;②剔除景点名称、用户 ID 和评论内容都相同的评论。经过以上处理,最终得到有效评论数据 57 122 条,数据样例如表 2 所示:

表2 用户评论实验数据样例

景点名称	用户ID	评论内容	评分	评论时间
湖北省博物馆	118 ****824	厚重的历史,馆内有不少精品,博物馆本身的建筑也很有楚文化的特色	5	2021-02-13
东湖风景区	小汐同学爱旅游	景色美,而且很多景点是免费的,比如行吟阁、亚洲棋院,都是很漂亮的古风建筑	5	2021-02-19
古琴台	m41 ****88	相传是俞伯牙和钟子期高山流水遇知音的地方,有天下知音第一台之称	5	2020-09-10
归元禅寺	木木	佛教圣地,文化底蕴深厚,环境不错。寺庙里面可以数罗汉,还可以求福	5	2020-05-15
木兰草原	M28 ****86	小火车很可爱,烟火秀很美,篝火晚会蕴含着丰富多彩的传统文化特色	5	2020-10-02
武昌起义纪念馆	mercury97	院外有孙中山先生的雕塑,院内前后有两座楼,前身是鄂军都督府,颇具气势	5	2018-11-21

4.2 标签生成结果

4.2.1 显式标签生成结果

显式标签包括物质文化标签、非物质文化标签和文化活动标签3类,依据3.2.2(1)所述方法进行实验,实验中各参数的取值为 $N_n = 1\ 000, N_v = 500, N_p = 500, T_n = 300, T_v = 500, T_p = 50, N = 200$。生成显式标签集合后,随机选取6个景点,使用词云工具WordCloud[32]进行可视化,各景点的标签词云图如图3所示:

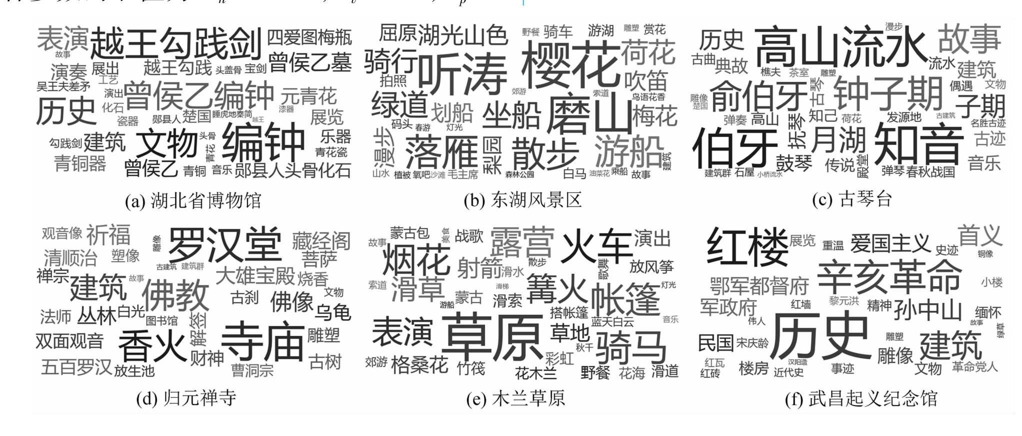


图3 景点的标签词云图

观察词云图可知,生成的显式标签能够较好地描述景点的文化资源特征。为进一步验证所提方法在显式标签生成上的性能,分别使用TF-IDF、TextRank和笔者所提出的方法提取关键词,并计算3种方法在抽取5-50个不同数量关键词时的准确率、召回率和F1值[33],实验结果见图4。

由图4可知,笔者所提方法的准确率、召回率和F1值都高于TF-IDF和TextRank,当抽取的关键词数量为50时,TF-IDF和TextRank的准确率、召回率和F1值普遍低于40%,且在不同景点上性能差异较大,在景点"东湖风景区"上3个指标都低于20%。本文方法的准确率、召回率和F1值普遍高于70%,且在不同景点上性能差异较小,具有更好的稳定性。此外,还需说明的是,虽然笔者所提出的方法在构建特征词表和噪声词表时需消耗一定人工,但实际上消耗的人工成本很低,构建针对武汉市57个景点的通用特征词表和噪声词表共计耗时约25分钟。在词表构建完成后,景点的显式标签集合便可通过算法自动生成。

生成显式标签集合后,按类别对标签进行分类,分别得到物质文化标签、非物质文化标签和文化活动标签,见表3。

4.2.2 隐式标签生成结果

隐式标签包括文化主题标签、文化感知强度标签和文化感知相似度标签3类。首先,依据3.2.2(2)所述方法,得到各景点的文化主题标签,见表4。

然后,依据3.2.2(3)所述方法,计算各景点的文化感知强度。文化感知强度包括3个指标,表5显示了整体文化感知强度和基于显式标签的文化感知强度的计算结果,表6显示了基于文化主题标签的文化感知强度计算结果。在计算整体文化感知强度时,各参数的取值为 $W_1 = 0.4, W_2 = 0.2, W_3 = 0.4$。强度值均经过归一化处理,数值范围在0 - 1之间。

由表5中整体文化感知强度计算结果可知,用户对寺庙和博物馆类型的景点普遍感知较强,在Top10中,寺庙有古德寺、归元禅寺和长春观,博物馆有江汉关博物馆、湖北省博物馆和辛亥革命博物馆。相比于

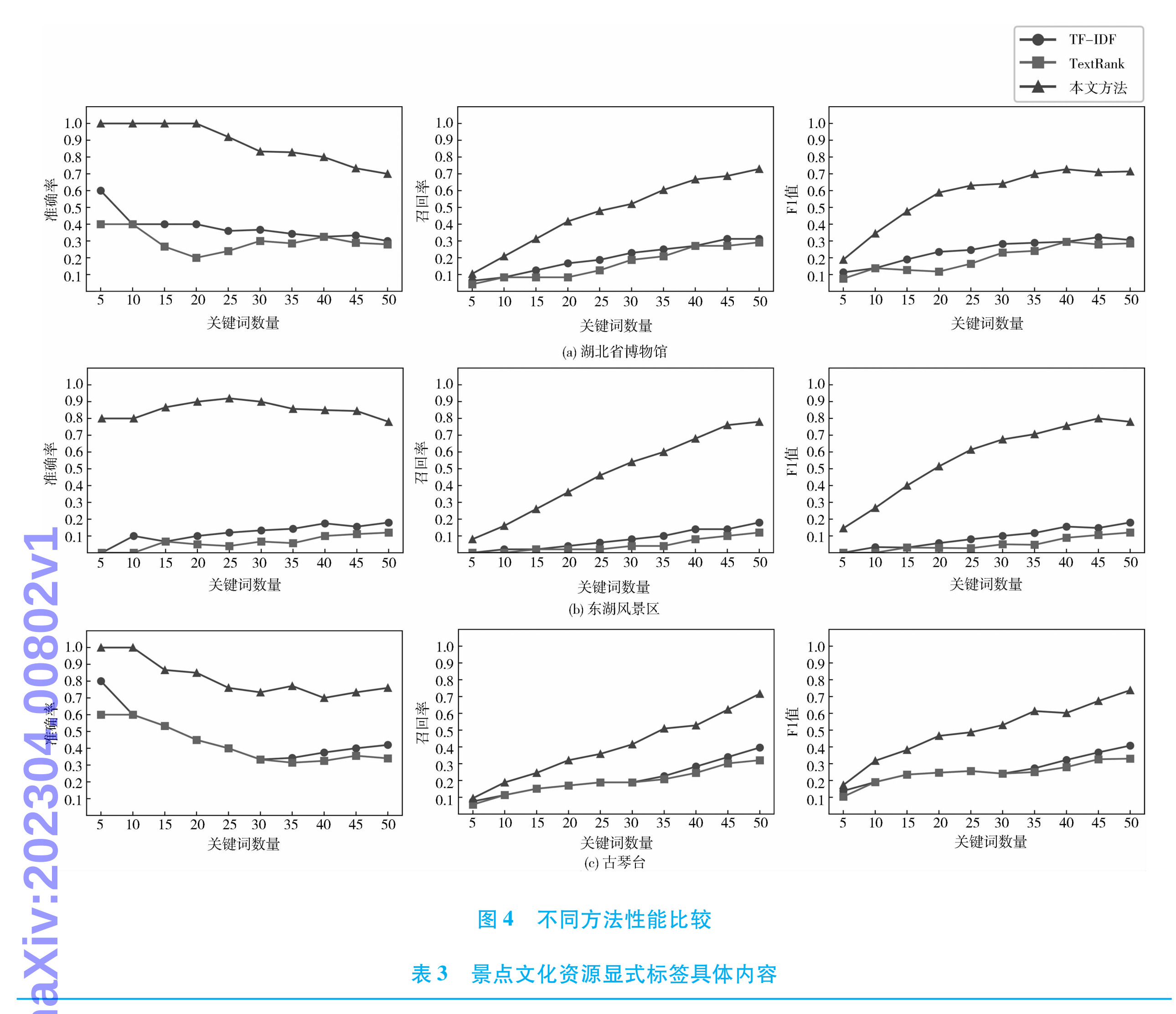


图 4 不同方法性能比较

表 3 景点文化资源显式标签具体内容

景点名称	物质文化标签	非物质文化标签	文化活动标签
湖北省博物馆	越王勾践剑 文物 曾侯乙编钟 建筑 青铜器 四爱图梅瓶 郧县人头骨化石 乐器 楚国 曾侯乙墓 吴王夫差矛 瓷器 青花瓷 睡虎地秦简 漆器	曾侯乙 越王勾践 音乐 工艺 故事 历史	表演 演奏
东湖风景区	樱花 落雁 绿道 荷花 湖光山色 梅花 梨园 码头 建筑 氧吧 山水 植被 鸟语花香 灯光 雕塑 沙滩 森林公园 油菜花 森林 花园 磨山 白马	屈原 毛主席 故事	游船 吹笛 骑车 赏花 乘船 春游 索道 郊游 帐篷 坐船 骑行 划船 漫步 游湖
古琴台	建筑 流水 文物 荷花 建筑群 雕像 楚国 雕塑 古建筑 小桥流水 月湖 古琴	知音 高山流水 伯牙 钟子期 俞伯牙 故事 音乐 子期 历史 鼓琴 抚琴 典故 知己 春秋战国 古曲 弹奏 弹琴 发源	春游 漫步
归元禅寺	罗汉 寺庙 罗汉堂 香火 寺院 建筑 佛像 丛林 大雄宝殿 观音 雕塑 图书馆 文物 雕像 建筑群 古建筑 藏经阁 乌龟 五百罗汉 菩萨 财神 双面观音 古树 法师 塑像 古刹 放生池 观音像	佛教 故事 清顺治 禅宗 曹洞宗	祈福 烧香 解签 礼佛
木兰草原	火车 草地 草坪 花海 美食 灯光 草原 格桑花 彩虹 蒙古包 蓝天白云	花木兰 音乐 故事 战歌 歌舞	骑马 烟花 篝火 滑草 射箭 放风筝 滑索 郊游 索道 秋千 游船 滑梯 露营 表演 演出 滑道 竹筏 搭帐篷 滑水
武昌起义纪念馆	红楼 建筑 雕像 文物 雕塑 鄂军都督府 军政府 楼房 小楼 红墙 红瓦 红砖 绿草 铜像	故事 历史 辛亥革命 首义 孙中山 爱国主义 精神 革命党人 史迹 事迹 宋庆龄 近代史 黎元洪 伟人	缅怀 重温

表 4　景点文化主题标签

景点	文化主题标签
湖北省博物馆	历史文化 楚文化
黄鹤楼	建筑文化 名人文化 历史文化
东湖	生态文化 楚文化
户部巷	建筑文化 历史文化 商贸文化 饮食文化
武汉长江大桥	建筑文化 塔桥文化
武汉大学	生态文化 建筑文化 历史文化 校史文化
木兰草原	生态文化 名人文化 民族文化
毛泽东同志旧居	红色文化 建筑文化 名人文化 历史文化
……	……
古琴台	建筑文化 名人文化 历史文化

其他景点，寺庙和博物馆的空间范围相对不大，但文化资源却十分丰富，寺庙有古建筑群和各式佛像，博物馆有大量文物，这些文化资源都密集地呈现在用户面前。此外，寺庙中不绝于耳的佛教音乐和念经声以及博物馆中用以辅助展示的灯光和设备，都使置身其中的用户被浓厚的文化氛围所感染，进而产生了更强的文化感知。

使用户产生较强物质文化感知的景点也主要是寺庙和博物馆两种类型，因其原因和整体文化感知类似，故不再赘述。结合非物质文化感知强度 Top10 中各景点的文化资源特征可知，计算结果与景点实际情况相符。例如，强度值最高的 3 个景点古琴台、辛亥革命博物馆和禹稷行宫的确都包含了广为人知的非物质文化资源，分别是“高山流水”的历史故事、“辛亥革命”的历史事件和“大禹治水”的神话传说。文化活动感知较强的景点基本都属于自然生态类型，Top10 景点中木兰天池、锦里沟、九真山、东湖等皆是如此。其原因是自然生态类景点通常依山傍水、占地面积广，景区管理者可因地制宜地开发骑马、射箭和篝火晚会等各类文化活动，进而丰富用户文化体验，增强文化感知。通过以上分析可知，各类文化感知强度计算结果与实际情况基本相符，同时也证明了所提方法的合理性和准确性。

表 5　整体文化感知强度和基于显式标签的文化感知强度计算结果

整体文化感知强度 Top10		物质文化感知强度 Top10		非物质文化感知强度 Top10		文化活动感知强度 Top10	
景点名称	强度值	景点名称	强度值	景点名称	强度值	景点名称	强度值
古德寺	1.0	古德寺	1.0	古琴台	1.0	木兰天池	1.0
归元禅寺	0.699 109	归元禅寺	0.766 107	辛亥革命博物馆	0.655 690	锦里沟	0.996 880
江汉关博物馆	0.620 274	湖北省博物馆	0.591 048	禹稷行宫	0.586 724	木兰草原	0.909 282
禹稷行宫	0.601 558	江汉关博物馆	0.579 189	古德寺	0.439 271	木兰湖	0.697 816
长春观	0.592 710	户部巷	0.533 525	毛泽东同志旧居	0.430 477	木兰清凉寨	0.647 800
湖北省博物馆	0.583 885	长春观	0.500 065	首义广场	0.424 389	九真山风景区	0.596 424
古琴台	0.579 850	武汉大学	0.475 560	长春观	0.423 461	东湖风景区	0.586 962
晴川阁	0.465 949	吉庆街	0.467 189	武昌起义纪念馆	0.389 667	东湖磨山景区	0.516 668
龟山公园	0.459 711	宝通禅寺	0.460 823	江汉关博物馆	0.300 230	汉口江滩公园	0.511 524
辛亥革命博物馆	0.458 839	汉阳造艺术区	0.448 900	龟山公园	0.221 162	海昌极地海洋公园	0.493 005

由表 6 中数据可知，同一景点不同文化主题的感知强度通常存在差异，例如宝通禅寺的宗教文化感知强于建筑文化感知，木兰湖的生态文化感知强于名人文化感知。不同景点在同一文化主题上的感知强度通常也会不同，因此在为用户提供基于文化主题的景点检索和推荐时，可按强度值大小对结果进行排序。

最后，依据 3.2.2(4) 所述方法，计算景点间的文化感知相似度，相似度用两特征向量的余弦值表示，数值范围是[－1，1]。表 7 以湖北省博物馆、东湖风景区、归元禅寺和武昌起义纪念馆为例，展示了与目标景点相似度最高和最低的部分景点。

从计算结果可以看出，相似度较高的景点通常都包含了 1－2 个相同的文化主题，例如东湖风景区、东湖磨山景区、珞珈山都包含“生态文化”，归元禅寺、宝通禅寺、古德寺都包含“建筑文化、宗教文化”。而相似度较低的景点通常没有相同的文化主题，且在文化资源内容上差异很大。例如，与湖北省博物馆相似度最低的景点是木兰天池，前者是室内的文物展览，后者是户外的生态旅游。与东湖风景区相似度最低的景点是杜莎夫人蜡像馆，前者是户外的生态旅游，后者是室内的蜡像展览。通过以上分析可知，景点间文化感知相似度的结果与实际情况基本相符。此外，为进一步评估隐式标签生成质量，笔者采用 5 级量表对上文中表 5、表 6、表 7 的结果进行专家评测，用“1 分”“2 分”“3 分”“4 分”“5 分”分别表示结果与实际情况“很不相符”“不相符”“一般”“相符”“很相符”，受邀被调查人员是 7 位对武汉市景点具有较高熟悉度的领域专家，根据调查结果统计得到“文化感知强度标签” 相关

表6 基于文化主题标签的文化感知强度计算结果

建筑文化 Top10		生态文化 Top10		宗教文化 Top5		名人文化 Top10	
景点名称	强度值	景点名称	强度值	景点名称	强度值	景点名称	强度值
古德寺	1.0	东湖樱花园	1.0	古德寺	1.0	古琴台	1.0
江汉关博物馆	0.868 892	武汉大学	0.982 813	归元禅寺	0.846 795	毛泽东同志旧居	0.967 992
汉口近代建筑群	0.674 719	木兰湖	0.824 497	长春观	0.779 312	禹稷行宫	0.747 222
禹稷行宫	0.594 276	中国地质大学逸夫博物馆	0.804 606	宝通禅寺	0.655 872	龟山公园	0.177 613
江汉路步行街	0.580 002	木兰云雾山	0.794 802	龟山公园	0.004 872	杜莎夫人蜡像馆	0.129 316
宝通禅寺	0.418 133	东湖磨山景区	0.719 060			武昌起义纪念馆	0.126 199
晴川阁	0.396 360	金龙水寨生态乐园	0.501 621			首义广场	0.106 525
大余湾	0.390 042	木兰天池	0.486 021			木兰湖	0.101 894
武汉大学	0.286 057	木兰清凉寨	0.480 483			珞珈山	0.076 420
武汉长江大桥	0.248 631	东湖风景区	0.434 849			中山公园	0.049 969

表7 景点间文化感知相似度计算结果

湖北省博物馆相似景点		东湖风景区相似景点		归元禅寺相似景点		武昌起义纪念馆相似景点	
景点名称	相似度	景点名称	相似度	景点名称	相似度	景点名称	相似度
武汉市博物馆	0.935 667	东湖磨山景区	0.755 117	宝通禅寺	0.916 938	辛亥革命博物馆	0.997 418
中华奇石馆	0.796 815	珞珈山	0.722 327	古德寺	0.913 906	毛泽东同志旧居	0.930 381
首义广场	0.730 050	木兰湖	0.695 691	长春观	0.852 768	首义广场	0.923 574
中国地质大学逸夫博物馆	0.729 626	张公山寨	0.692 782	禹稷行宫	0.844 981	江汉关博物馆	0.908 176
湖北省美术馆	0.728 325	东湖樱花园	0.681 611	汉阳造艺术区	0.825 852	汉口近代建筑群	0.901 440
……	……	……	……	……	……	……	……
木兰古门	0.131 494	武昌起义纪念馆	0.098 400	木兰清凉寨	-0.003 026	紫薇都市田园	0.023 875
木兰清凉寨	0.013 309	归元禅寺	0.084 988	九真山风景区	-0.028 534	金龙水寨生态乐园	-0.032 508
九真山风景区	-0.066 678	户部巷	0.057 286	木兰草原	-0.056 063	木兰清凉寨	-0.043 515
木兰天池	-0.121 245	杜莎夫人蜡像馆	-0.049 495	木兰天池	-0.132 850	木兰天池	-0.093 649

结果的平均得分是4.375分，“文化感知相似度标签”相关结果的平均得分是4.357分，由此表明笔者所提出的隐式标签生成方法具有较高的可靠性。

4.3 标签展示

生成显式标签和隐式标签后，每一个景点都既可以使用标签云构建文化资源画像（见图3），也可以通过表格对标签进行详细展示。表8以湖北省博物馆为例，展示了生成的各类标签。其中，标签类别后的数字表示该类标签的文化感知强度，数值范围是[1, 10]，由前文中归一化后的数值乘以10得到。由表8中数据可知，湖北省博物馆的整体文化感知强度是5.84，历史文化和楚文化的感知强度都是10.0，物质文化、非物质文化和文化活动的感知强度分别是5.91、1.19和4.32。用户可根据各类标签的具体内容和感知强度判断景点的文化资源是否与自己的偏好相符。

表8 湖北省博物馆文化资源标签展示

景点名称	文化主题	物质文化(5.91)	非物质文化(1.19)	文化活动(4.32)
湖北省博物馆(5.84)	历史文化(10.0) 楚文化(10.0)	越王勾践剑 文物 曾侯乙编钟 建筑 青铜器 四爱图梅瓶 郧县人头骨化石 乐器 楚国 曾侯乙墓 吴王夫差矛 瓷器 青花瓷 睡虎地秦简 漆器	曾侯乙 越王勾践 音乐 工艺 故事 历史	表演 演奏

4.4 标签应用

利用生成的文化资源标签，可根据用户的需求和偏好，提供更加精准的检索和推荐服务。下面对标签在景点检索和推荐两个方面的应用进行分析和示例。

4.4.1 景点检索

当用户有较为明确的兴趣对象时，可以通过关键词进行景点检索。例如，用户听闻武汉的樱花很有名，想知道武汉有哪些景点可以观赏樱花，此时便可以使用关键词“樱花”进行检索。

这里存在两种情况：一种是用户输入的关键词有与之匹配的标签，另一种是关键词与所有标签都不匹配。针对以上两种情况，笔者提供了基于标签内容的检索和基于标签语义的检索两种方式。

（1）基于标签内容的检索。适用场景：存在与检

索的关键词相匹配的标签。

检索方法：根据用户输入的关键词，检索出所有标签中含有该关键词的景点，并对景点按整体文化感知强度排序后展示给用户。

检索结果：以关键词"樱花"为例，根据以上方法得到的 Top5 景点为珞珈山、武汉大学、东湖樱花园、昙华林、东湖磨山景区。

(2) 基于标签语义的检索。适用场景：不存在与检索的关键词相匹配的标签。

检索方法：首先通过已训练好的词向量模型，获取关键词的词向量，然后计算关键词与景点所有显式标签的余弦相似度的平均值，并将景点按相似度大小排序后展示给用户。

检索结果：假如用户对古典建筑感兴趣，且以关键词"古典建筑"进行检索，此时会发现标签中只有"古建筑"，没有"古典建筑"。根据上述基于标签语义的检索方法可得到 Top5 景点列表是：晴川阁、长春观、归元禅寺、古德寺、黄鹤楼。本文方法除给出检索结果外，还可对结果进行解释。相关研究表明，将检索和推荐的具体原因展示给用户，可显著提升用户对结果的信任度[34－35]。表 9 列举了检索结果的 Top5 景点，以及对结果的解释。

表 9　关键词"古典建筑"基于标签语义的检索结果

检索结果 Top5	对结果的解释
景点名称	该景点包含的以下标签与关键词相似
晴川阁	古建筑 建筑 建筑群 楼阁 铁门关
长春观	古建筑 建筑 建筑群 道观 殿堂
归元禅寺	古建筑 建筑 建筑群 大雄宝殿 藏经阁
古德寺	建筑 宗教建筑 宝殿 神庙
黄鹤楼	建筑 宝塔 牌坊

4.4.2　景点推荐

用户信息一般包括性别、年龄等个人基本信息，以及其在旅游网站上的检索记录和游览历史。根据以上信息分析出用户的兴趣偏好后，便可基于文化资源标签进行个性化推荐。针对不同的服务场景，笔者提供了 3 种推荐方式：

(1) 基于文化主题标签的推荐。适用场景：已根据用户的检索记录或游览历史，获知了用户的文化主题偏好。

推荐方法：首先，根据"文化主题—景点"矩阵，得到用户感兴趣的文化主题下的所有景点；然后，既可以直接根据基于主题标签的文化感知强度的大小生成推荐列表(见表 6)，也可以综合文化感知强度、景点热度和景点好评度等多种因素，通过指标加权生成最终推荐列表。

推荐结果：以用户偏好的文化主题是"生态文化"为例，表 10 展示了基于不同因素的景点推荐结果。对比结果可知，当仅考虑文化感知强度时，木兰湖、东湖磨山景区、金龙水寨生态乐园等景点都在 Top10 中，且排名靠前。当综合热度或好评度等因素后，以上景点的排名出现明显下降。实际上，如果仅考虑文化感知强度，部分冷门的景点可能会获得较高排名，例如木兰湖。当加入热度或好评度后，可以通过对权值的调整有效剔除部分过于冷门或好评度低的景点，使推荐结果能更好地满足用户的实际需求。

表 10　"生态文化"主题景点推荐结果对比

仅考虑文化感知强度的 Top10 推荐结果	文化感知强度＋热度的 Top10 推荐结果	文化感知强度＋好评度的 Top10 推荐结果
东湖樱花园	海昌极地海洋公园	东湖樱花园
武汉大学	东湖海洋世界	武汉大学
木兰湖	武汉大学	中国地质大学逸夫博物馆
中国地质大学逸夫博物馆	东湖樱花园	木兰天池
木兰云雾山	中国地质大学逸夫博物馆	木兰云雾山
东湖磨山景区	木兰天池	珞珈山
金龙水寨生态乐园	木兰云雾山	东湖风景区
木兰天池	木兰草原	木兰清凉寨
木兰清凉寨	木兰湖	武汉植物园
东湖风景区	武汉植物园	汉口江滩公园

(2) 基于显式标签的推荐。适用场景：已根据用户的检索记录或游览历史，获知了用户的文化类别偏好。

推荐方法：以用户偏好是"文化活动"为例，首先根据"文化活动—景点"矩阵，得到按文化活动感知强度排序的景点列表(见表 4)；然后根据 4.4.2(1) 小节所述方法生成推荐列表。

推荐结果：因推荐方法与 4.4.2(1) 小节相同，故不再具体展示。

(3) 基于文化感知相似度标签的推荐。适用场景：当已知用户对某个景点感兴趣，或用户已经从提供的推荐列表中选定某个景点时，可继续向用户推荐与该景点相似的其他景点。

推荐方法：既可以按相似度大小排序生成推荐列表(见表 7)，也可以综合景点热度、景点好评度等因素，通过指标加权生成推荐列表。

推荐结果：以用户选定的景点是"古德寺"为例，表 11 展示了基于文化感知相似度＋景点热度＋景点好评度的 Top10 推荐结果。对推荐结果的解释一直是

个性化推荐研究中的重点和难点问题，本文方法在给出推荐结果的同时，也可展示给用户每一个景点的具体推荐理由，这有助于提升用户对结果的接受度和信任度。

表 11 “古德寺”相似景点推荐结果和推荐理由

文化感知相似度 + 热度 + 好评度的 Top10 推荐结果	推荐理由	
	相同的文化主题	相同的标签
汉口近代建筑群	建筑文化	建筑 历史
江汉关博物馆	建筑文化	建筑 历史
归元禅寺	建筑文化 宗教文化	寺庙 香火 建筑 佛像 大雄宝殿 佛教 祈福
江汉路步行街	建筑文化	建筑 历史
宝通禅寺	建筑文化 宗教文化	寺庙 建筑 大雄宝殿 佛像 历史 佛教 祈福
大余湾	建筑文化	建筑 荷花 历史
黄鹤楼	建筑文化	建筑
禹稷行宫	建筑文化	建筑 寺庙 神庙 历史
长春观	建筑文化 宗教文化	建筑 香火 寺庙 道教 佛教 宗教 历史
昙华林	建筑文化 宗教文化	建筑 教堂 历史

5 结语

笔者针对文化旅游信息检索困难、推荐形式单一的问题，提出了一种基于在线旅游信息的景点文化资源标签自动生成方法，同时针对信息服务中的不同场景，提供了基于标签内容和基于标签语义的两种检索方法，以及基于文化主题标签、基于显式标签和基于文化感知相似度标签的 3 种推荐方法。并以武汉市的景点为例，利用真实数据集验证了所提方法的可行性。

本文的主要贡献包括：①提出从文化资源视角对景点进行描述、检索和推荐，设计了文化资源标签体系，为旅游景点的组织管理提供了新的思路；②提出一种基于特征词筛选和噪声词过滤的文化资源显式标签生成方法，实验结果表明，相比于传统方法，该方法具有更好的性能；③设计文化感知强度和文化感知相似度的计算方法，实验表明，文化感知强度和文化感知相似度能够有效反映不同景点的文化特征差异和用户对文化资源感知的实际情况，同时也是景点检索结果排序和景点推荐的重要依据；④基于生成的标签，提供了两种检索方法和 3 种推荐方法，且所提方法均具备可解释性，有效提升了检索和推荐的透明度以及用户对结果的信任度，同时也为其他领域的推荐解释性研究提供了参考。

本文的不足之处在于：生成显式标签时，虽然通过算法可自动生成显式标签集合，但仍需人工分类才能得到不同类别的标签。通过指标加权生成推荐列表时，指标的权值仅使用了经验值，在实际应用中应根据用户偏好进行动态设置。此外，对隐式标签生成结果进行评估时，参与评估的专家数较少也是本文的不足之处。今后的工作将从 3 个方面展开：①研究如何实现显式标签的自动分类；②指标权值会对推荐结果产生重大影响，应将用户偏好与权值关联起来，研究基于用户偏好的权值动态设置方法，使推荐结果更加符合用户的实际需求；③研究如何对基于文化资源的景点检索和推荐结果进行有效的量化评估。

参考文献：

[1] 张朝枝，朱敏敏. 文化和旅游融合：多层次关系内涵、挑战与践行路径[J]. 旅游学刊，2020，35(3)：62－71.
[2] JELASSI M N，YAHIA S B，NGUIFO E M. Towards more targeted recommendations in folksonomies[J]. Social network analysis and mining，2015，5(1)：1－18.
[3] GUPTA M，LI R，YIN Z，et al. Survey on social tagging techniques[J]. ACM sigkdd explorations newsletter，2010，12(1)：58－72.
[4] 李旭晖，李媛媛，马费成. 我国图情领域社会化标签研究主要问题分析[J]. 图书情报工作，2018，62(16)：120－131.
[5] 孟秋晴，熊回香，童兆莉，等. 基于在线问诊文本信息的医生标签自动生成研究[J]. 情报科学，2020，38(5)：58－64,72.
[6] SHEN Z，ARSLAN AY S，KIM S H，et al. Automatic tag generation and ranking for sensor-rich outdoor videos[C]// Proceedings of the 19th ACM international conference on multimedia. New York：Association for Computing Machinery，2011：93－102.
[7] 叶佳鑫，熊回香，童兆莉，等. 在线医疗社区中面向医生的协同标注研究[J]. 数据分析与知识发现，2020，4(6)：118－128.
[8] 吴小兰，章成志. 结合用户关系网和标签共现网的微博用户标签推荐研究[J]. 情报学报，2015，34(5)：459－465.
[9] 熊回香，叶佳鑫. 基于 LDA 主题模型的微博标签生成研究[J]. 情报科学，2018，36(10)：7－12.
[10] 蒋武轩，易明，熊回香，等. 网络社交平台中社群标签生成研究[J]. 图书情报工作，2021，65(10)：79－89.
[11] ZENG L，GUO X，YANG C，et al. TagNN：a code tag generation technology for resource retrieval from open-source big data[J]. Wireless communications and mobile computing，2021：9956207.
[12] 赵辉，化柏林，何鸿魏. 科技情报用户画像标签生成与推荐[J]. 情报学报，2020，39(11)：1214－1222.
[13] 李蕾，章成志. 社会化标签质量评估研究综述[J]. 现代图书情报技术，2013(11)：22－29.
[14] 章成志，赵华，李蕾，等. 中英文图片标签质量差异比较研究——以 Flickr 为例[J]. 情报理论与实践，2018，41(4)：123－127.
[15] 章成志，李蕾. 社会化标签质量自动评估研究[J]. 现代图书情报技术，2015(10)：2－12.

[16] 章成志，顾晓雪. 区分标签质量的机器生成标签聚类研究[J]. 现代图书情报技术，2015(10)：22－29.
[17] 李雅美，王昌栋. 基于标签的个性化旅游推荐[J]. 中国科学技术大学学报，2017，47(7)：547－555.
[18] 史一帆，文益民，蔡国永，等. 基于景点标签的协同过滤推荐[J]. 计算机应用，2014，34(10)：2854－2858.
[19] 单晓红，张晓月，刘晓燕. 基于在线评论的用户画像研究——以携程酒店为例[J]. 情报理论与实践，2018，41(4)：99－104,149.
[20] 刘海鸥，孙晶晶，苏妍嫄，等. 基于用户画像的旅游情境化推荐服务研究[J]. 情报理论与实践，2018，41(10)：87－92.
[21] 赵平，孙连英，涂帅，等. 改进的知识迁移景点实体识别算法研究及应用[J]. 数据分析与知识发现，2020，4(5)：118－125.
[22] 刘小安，彭涛. 基于卷积神经网络的中文景点识别研究[J]. 计算机工程与应用，2020，56(4)：140－145.
[23] 杨一帆，陈文亮. 旅游场景下的实体别名抽取联合模型[J]. 中文信息学报，2020，34(6)：55－63.
[24] 尹小娜，郑向敏. 基于网络文本分析的三坊七巷游客文化感知研究[J]. 北京第二外国语学院学报，2015，37(9)：62－66,33.
[25] 李东晔，黄震方，叶滨鸿，等. 游客慢文化感知维度分异与影响因素研究——以高淳国际慢城为例[J]. 人文地理，2020，35(1)：150－160.
[26] 庞朴. 文化结构与近代中国[J]. 中国社会科学，1986(5)：81－98.
[27] SALTON G，BUCKLEY C. Term-weighting approaches in automatic text retrieval[J]. Information processing & management，1988，24(5)：513－523.
[28] MIHALCEA R，TARAU P. Textrank：bringing order into text[C]//Proceedings of the 2004 conference on empirical methods in natural language processing. Barcelona：Association for Computational Linguistics，2004：404－411.
[29] SUN J. "Jieba" Chinese text segmentation：built to be the best Python Chinese word segmentation module[EB/OL]. [2022－01－15]. https://github. com/fxsjy/jieba.
[30] BENGIO Y，DUCHARME R，VINCENT P，et al. A neural probabilistic language model[J]. Journal of machine learning research，2003，3(Feb)：1137－1155.
[31] MIKOLOV T，CHEN K，CORRADO G，et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv：1301.3781，2013.
[32] ANDREAS M. Word_cloud[EB/OL]. [2022－01－15]. https://github. com/amueller/word_cloud.
[33] 赵京胜，朱巧明，周国栋，等. 自动关键词抽取研究综述[J]. 软件学报，2017，28(9)：2431－2449.
[34] 项亮. 推荐系统实践[M]. 北京：人民邮电出版社，2012.
[35] 王喆. 深度学习推荐系统[M]. 北京：电子工业出版社，2020.

作者贡献说明：
郑淞尹:设计研究方案,数据采集与实验分析,论文撰写与修改;
谈国新:提出研究思路,论文审阅与修改。

Research on Automatic Generation and Application of Cultural Resource Tags of Scenic Spots

Zheng Songyin　Tan Guoxin
National Research Center of Cultural Industries，Central China Normal University，Wuhan 430079

Abstract：[**Purpose/Significance**] To generate high-quality cultural resource tags for scenic spots，and solve the problems of difficult information retrieval and signal recommendation form in cultural tourism services. [**Method/Process**] First，a tag system for cultural resources including explicit and implicit tag types was designed；then，an explicit tag generation method based on feature word filtering and noise word filtering was proposed，and the calculation method of cultural perception intensity and cultural perception similarity in implicit tags was designed，and cultural resource tags of scenic spots were generated based on the above methods；finally，for different scenarios in tourism information services，retrieval methods and recommendation methods based on cultural resource tags were provided. [**Result/Conclusion**] Taking the real tourism data of Wuhan as an example to conduct empirical research. The results show that the tags generated based on this method can accurately describe the cultural resource characteristics of scenic spots，and the retrieval and recommendation algorithms based on tags have strong interpretability，which can effectively improve the transparency of information services and users' trust in the results，and have reference value for recommendation and interpretation research in other fields.

Keywords：tourism information service　tag generation　resource retrieval　travel recommendation